

Satisfaction relations for proper classes: Applications in logic and set theory

Robert A. Van Wesep

*1402 Bolton Street
Baltimore, MD, 21217, USA*

Abstract

We develop the theory of partial satisfaction relations for structures that may be proper classes and define a satisfaction predicate (\models^*) appropriate to such structures. We indicate the utility of this theory as a framework for the development of the metatheory of first-order predicate logic and set theory, and we use it to prove that for any recursively enumerable extension Θ of ZF there is a finitely axiomatizable extension Θ' of GB that is a conservative extension of Θ . We also prove a conservative extension result that justifies the use of \models^* to characterize ground models for forcing constructions.

Keywords: satisfaction, proper class, class theory, conservative extension

email: rvanwesep@verizon.net

1 Introduction

In discussions of the theory of sets it is a common practice to use the terminology of satisfaction and models informally with reference to structures that are proper classes. An “inner model”, for example, is informally defined as a proper transitive class M such that $(M; \in)$ “satisfies” ZF (Zermelo-Fraenkel set theory). In the context of a pure set theory, such as ZF, any such reference is necessarily informal for two reasons:

1. Proper classes do not exist, and can only be referred to as predicates applicable to sets.
2. ZF is not finitely axiomatizable, so one cannot say that M is an inner model by means of a single sentence relativized to M .

For example, in the context of ZF the statement

$$(1.1) \quad L \text{ is a model of ZFC}$$

is understood to stand for the set of all sentences θ^L , where θ is an axiom of ZFC, and θ^L is the sentence θ with all quantified variables are restricted to constructible sets. In the context of a class theory such as the von Neumann-Bernays-Gödel, or Gödel-Bernays, theory GB, one can demonstrate the existence

of the class L of constructible sets, and one might expect to be able to formulate (1.1) as a single sentence

$$(1.2) \quad L \models \text{ZFC},$$

but the interpretation of such a sentence is problematic, inasmuch as it is not possible to prove—in a conservative¹ class theory such as GB—that there exists a satisfaction relation for L (or for proper class structures in general).

In this article, we examine issues related to the existence and use of partial satisfaction relations for proper class structures in the context of conservative class theories with and without the Infinity axiom. In particular, we propose a definition of satisfaction for proper class structures that is weak enough that (1.2), for example, is a theorem of GB, but strong enough that (1.2) implies θ^L for every axiom θ of ZF. For the sake of emphasis, in this article we denote this *universal satisfaction predicate* by ‘ \models^* ’, but there is no reason its definition (2.10) could not be taken as the primary definition of ‘ \models ’. By means of \models^* , the conventional informal use of proper class models may be rendered formally correct with minimal modifications to standard practice. Such notions as elementary substructure and elementary embedding of proper class structures are formalizable in GB in a similar way, as are the notions of forcing relation and boolean valuation, all of which are subject to the same limitations as satisfaction in conventional treatments.

Section 2 is devoted to a brief explication of the notion of satisfaction for proper class structures and related ideas. These are quite straightforward, and we do not suppose that they are entirely new; however, we have not found any systematic treatment of them in the existing literature. In particular, we have not found a definition of \models^* or a discussion of its use in set theory. We therefore state the principal theorems and sketch their (straightforward) proofs. We do this primarily in the context of the theories S and C , which we define respectively to be ZF and GB with the Infinity axiom omitted. Note that these theories do not contain \neg Infinity, so they may be extended to ZF and GB, respectively, by the addition of Infinity. $S + \neg$ Infinity is—for all practical purposes—Peano arithmetic. The use of satisfaction relations for proper classes in C permits an essentially finitary, but nonetheless efficient, development of the theory of first-order predicate logic.² The book *Foundations of Mathematics: A Generalist's Guide*[10] employs these ideas throughout and demonstrates the clarity and

¹By ‘conservative’ we mean that the class comprehension axiom schema employs only formulas without bound class variables.

²We do not present this development here beyond what is necessary for the purposes of this article; but the following example is worthy of mention. Consider Gödel’s first incompleteness theorem, which presents a sentence σ that says, in effect, that σ is not provable in S (which we use here for convenience instead of the more customary Peano arithmetic), and states that σ is not a theorem of S , even though it is (by virtue of that fact) true. This is ordinarily proved in a metatheory other than S , which is often not explicitly characterized, but may generally be considered to be ZF. Then it is stated that the theorem and proof could be given in S with the additional hypothesis that S is consistent. Therefore, assuming $\text{Con } S$, $S \not\vdash \text{Con } S$, which is Gödel’s second incompleteness theorem. An actual proof of the first theorem in S is generally considered such an ungainly thing that it is not usually given in any detail, despite the fact that its existence is critical to the proof of the second theorem. This defect can be remedied

precision they bring to the explication of the foundations of mathematics, from elementary logic to advanced set theory.

In Section 3 we prove a generalization of the well known theorem that GB is a finitely axiomatizable conservative extension of ZF. This is Theorem 3.1. The universal satisfaction predicate is not involved in the statement of this result, but (a slight variation of) it is intrinsic to the proof.

In Section 4 we discuss the implications of the use of \models^* for the correlation between the forcing relation within a transitive model M of ZF, and the satisfaction relation in a generic extension $M[G]$. If M is a countable set, of course, $p \Vdash^{\mathbb{P}} \phi(x_0, \dots, x_{n-1})$ iff for every M -generic filter G on a partial order $\mathbb{P} \in M$, if $p \in G$ then $M[G] \models \phi[x_0^G, \dots, x_{n-1}^G]$. Essentially the same correlation holds in the general case (when M is an uncountable set or a proper class), but it must be stated without reference to generic filters.

This is the method of “arguing in a generic extension”, which conventionally goes as follows. The theory Θ (Definition 4.1)—which includes ZF and also says that the universe is $V[G]$, where V is an inner model of ZF, and G is a V -generic filter on \mathbb{P} —holds in $M[G]$ whenever G is an M -generic filter on $\mathbb{P} \in M$, and V is interpreted as M . “Arguing in a generic extension” means proving something in Θ to demonstrate that something is forced.

If we adopt \models^* , it is natural to use the theory Θ' defined in (4.7), which differs from Θ in that it includes GB and implements the assumption that V is an inner model of ZF as the single sentence $\ulcorner V \models^* \text{ZF} \urcorner$, rather than all sentences $\ulcorner \theta^V \urcorner$, where θ is an axiom of ZF. To justify the inference that something is forced from the existence of a proof in Θ' , we show that Θ' is a conservative extension of Θ . This is Theorem 4.9.

In Section 5, for the sake of completeness, we define a universal forcing relation and universal boolean valuation function that are analogous to the universal satisfaction relation, by means of which the conventional informal use of forcing relations and boolean valuations for proper classes may be rendered formally correct—again, with minimal modifications to standard practice.

2 The role of satisfaction for proper classes in set theory and its metatheory

2.1 Some useful conventions

Recall the definition of S and C as ZF and GB with Infinity omitted. We regard C (like GB) as a theory with two sorts of individuals: sets and classes. All sets are classes. A class is *proper* $\stackrel{\text{def}}{\iff}$ it is not a set. Let \mathbf{c} be the signature (similarity type) of this theory, and let \mathbf{s} be this signature with the class sort omitted. Thus, \mathbf{s} is the signature of S (and ZF). Let $\mathcal{L}^{\mathbf{s}}$ and $\mathcal{L}^{\mathbf{c}}$ be the languages

by proving the first theorem in C , rather than ZF, which is not any harder, and is arguably more natural. Then we can invoke the well known result that C is a conservative extension of S (2.2) to conclude rigorously that a proof of the first theorem in S exists, without giving it in detail.

appropriate to the above signatures. Given a theory Θ , i.e., a class of sentences, let $\overline{\Theta} \stackrel{\text{def}}{=} \text{its deductive closure}$.

It is well known that

(2.1) C is finitely axiomatizable,

whereas S is not.³

(2.2) C is a conservative extension of S in the sense that $\overline{C} \cap \mathcal{L}^S = \overline{S}$.

(2.2) has a simple infinitary proof (e.g., a ZF-proof), and a considerably more involved finitary proof, (e.g., an S-proof). Given (2.2), it is appropriate to regard C as a *finitary* theory, like S .

The discussion that follows, particularly the proofs of Theorems 3.1 and 4.9, involves a sufficiently intricate interplay of meta- and object theories that it is helpful to use notation for linguistic expressions that distinguishes use and mention more particularly than is often done; although the reader is forewarned that we do not always maintain the highest standard in this regard—a judicious ambiguity sometimes best serves the cause of clarity. Unless otherwise noted, the following discussion takes place in the context of C . Thus, infinite sets may not exist.

Given a text string that represents an expression ϵ in an object language, if we flank it with corner quotes, $\ulcorner \dots \urcorner$, we create a string that represents a metalanguage name ν for ϵ . We use boldface versions of standard typographic symbols for syntactical operations to denote various expression-building operations in any language. Thus, for example, $\phi \wedge \psi$ is the conjunction of formulas ϕ and ψ in any language. We may extend this to some common predicate and operation symbols, such as those for membership and identity. Thus, if $u = \ulcorner x \urcorner$, $v = \ulcorner y \urcorner$, and $w = \ulcorner z \urcorner$, then $u \in v \wedge v \in w \rightarrow u \in w$ is ‘if $x \in y$ and $y \in z$ then $x \in z$ ’.

Substitution of terms for variables in expressions is indicated with round brackets. Suppose ϵ is an expression, and v_0, \dots, v_{n-1} are in $\text{Free } \epsilon$, the set of free variables of ϵ . Suppose $\tau_0, \dots, \tau_{n-1}$ are terms. $\epsilon \left(\begin{smallmatrix} v_0 & \dots & v_{n-1} \\ \tau_0 & \dots & \tau_{n-1} \end{smallmatrix} \right)$ is the expression that results from the indicated substitutions. When it is not necessary to indicate the variables, ‘ $\epsilon(\tau_0, \dots, \tau_{n-1})$ ’ may be used. It is often convenient to use

³This is better known for GB and ZF. The finite axiomatization of GB as originally given by Bernays[1] yields C if Infinity is omitted. A proof that ZF is not finitely axiomatizable may be given in $S + \text{Con ZF}$ as an easy application of the reflection theorem schema of ZF, i.e. of the theorem of S that every instance of the reflection schema is a theorem of ZF. From this it follows that S is not finitely axiomatizable (otherwise by adding Infinity we would have a finite axiomatization of ZF); however, this gives the result as a theorem of $S + \text{Con ZF}$. In what may be viewed as a clever adaptation of the reflection method, Ryll-Nardzewski gave a proof in ZF that S is not finitely axiomatizable[6]. Ryll-Nardzewski’s argument is intrinsically infinitary, as it uses the satisfaction relation for $(V_\omega; \in)$. (In fact, the statement of his theorem is intrinsically infinitary, as it applies to arbitrary true extensions of S , i.e., theories $\Theta \supseteq S$ such that $\models^S \Theta$, where S is the full satisfaction relation for $(V_\omega; \in)$.) By way of illustrating some of the ideas presented in this article, we show in the proof of Theorem 2.25 how Ryll-Nardzewski’s argument may be (easily) adapted to provide a proof in $S + \text{Con } S$ that S is not finitely axiomatizable.

a similar notation where ϵ is indicated using the corner-quote convention. For example, suppose ϕ , ψ , and θ are respectively $\ulcorner x \in y \urcorner$, $\ulcorner y \in z \urcorner$, and $\ulcorner x \in z \urcorner$. Then

$$\ulcorner \text{if } (\phi) \text{ and } (\psi) \text{ then } (\theta) \urcorner \quad (2.3)$$

is

$$\ulcorner \text{if } x \in y \text{ and } y \in z \text{ then } x \in z \urcorner,$$

the result of substituting the metalanguage terms

$$\ulcorner x \in y \urcorner, \ulcorner y \in z \urcorner, \ulcorner x \in z \urcorner$$

in the metalanguage term

$$\ulcorner \text{if } _ \text{ and } _ \text{ then } _ \urcorner,$$

where the underscores indicate variables (ranging over formulas) that we do not need to name, as they are always substituted in the indicated fashion. When we indicate such a substitution “in line”, as in (2.3), we give the brackets a lighter tone than the surrounding text, so as to render them relatively unobtrusive and to distinguish them from round brackets used as grouping indicators.

A similar convention applies to the use of square brackets to indicate assignments of individuals in a structure to variables in an expression for the purpose of valuation (or satisfaction in the case of formulas). Our first use of this is in (3.6). Note that (3.6.1) is a conventional use of square brackets to create the statement that the \mathbf{c} -formula D is satisfied at the indicated values, viz., θ and y , for its free variables. (3.6.2) is the statement that the formula indicated by the corner-quoted text, with implicit variables in place of the insertions, is satisfied when those variables are assigned the indicated values, viz., S and n .

For convenience we suppose that \mathbf{s} and \mathbf{c} have two binary predicate symbols, one to denote membership and one to denote identity. The extended signatures \mathbf{s}' and \mathbf{c}' have, in addition, a nullary operation symbol (i.e., a constant) to denote the empty set 0 , and a binary operation symbol to denote the *add* operation: $x \frown y = x \cup \{y\}$.

‘HF’ is a defined predicate in \mathbf{S} characterizing the hereditarily finite sets. In \mathbf{C} , ‘HF’ may be used this way, and also as a constant denoting the class of hereditarily finite sets. Note that in this setting, $\text{HF} = V_\omega$. For each x such that $\text{HF } x$, let \hat{x} be a specific \mathbf{s}' -term (a composition of $\ulcorner 0 \urcorner$ and $\ulcorner \frown \urcorner$) whose value is x , chosen by some fixed recursive procedure. Call \hat{x} the *canonical name* of x .

We will use an informal representation of structures, such that

$$(D_0, D_1, \dots; X_0, X_1, \dots)$$

is a structure with domains (sorts) D_0, D_1, \dots ; and predicates and operations X_0, X_1, \dots . We regard $(D_0, D_1, \dots; X_0, X_1, \dots)$ as encoding D_0, \dots, X_0, \dots in a way that is applicable to proper classes, as well as sets.

We suppose that for any signature ρ , the expression-building operations for the corresponding language \mathcal{L}^ρ are uniformly defined in terms of 0 and \frown in such a way that the rank of any expression is greater than the rank of any of its subexpressions. If $\rho \in \text{HF}$ then $\mathcal{L}^\rho \subseteq \text{HF}$.

2.2 Partial satisfaction relations

Suppose ϵ is an expression. Then $\text{Free } \epsilon \stackrel{\text{def}}{=} \text{the set of free variables of } \epsilon$. A is an \mathfrak{S} -assignment for $\epsilon \stackrel{\text{def}}{\iff} A$ is a finite function into $|\mathfrak{S}|$ such that $\text{Free } \epsilon \subseteq \text{Dom } A$. The value of an expression ϵ at an \mathfrak{S} -assignment A for ϵ is an element of $|\mathfrak{S}|$ if ϵ is a term and is a member of 2, i.e., $\{0, 1\}$, if ϵ is a formula, with 1 corresponding to ‘true’ and 0 to ‘false’.

Suppose Φ is a class of ρ -expressions, $\overline{\Phi} \stackrel{\text{def}}{=} \text{the class of subexpressions of } \Phi$. We regard an expression as a subexpression of itself, so $\overline{\Phi} \supseteq \Phi$. Suppose \mathfrak{S} is a ρ -structure. A Φ -valuation function for \mathfrak{S} is a function F such that

1. $\text{Dom } F$ consists of all $\langle \epsilon, A \rangle$ such that $\epsilon \in \overline{\Phi}$ and A is an \mathfrak{S} -assignment for ϵ ;
2. $F\langle \epsilon, A \rangle$ is in $|\mathfrak{S}|$ if ϵ is a term and in 2 if ϵ is a formula;
3. F satisfies the usual recursive definition of valuation.

F is a *partial* valuation function for $\mathfrak{S} \stackrel{\text{def}}{\iff} F$ is a Φ -valuation for some class Φ of ρ -expressions.

Suppose \mathfrak{S} is a ρ -structure. It is straightforward to show in \mathbf{C} that partial valuations for \mathfrak{S} agree on their common domain.

The following theorem is essentially trivial, but we take the time to state it and sketch the proof to point out where the corresponding proof for the class \mathcal{F}^ρ of all ρ -formulas fails when \mathfrak{S} is a proper class. Here, as elsewhere, we indicate the theory within which a theorem is stated and proved; in this case it is \mathbf{C} . We do the same for definitions.

Theorem 2.4. [C] *Suppose \mathfrak{S} is a ρ -structure. Then there is a unique \mathcal{T}^ρ -valuation function for \mathfrak{S} , where \mathcal{T}^ρ is the class of ρ -terms.*

Proof. Suppose τ is a term and A is an assignment for τ . F is a $\langle \tau, A \rangle$ -valuation function for $\mathfrak{S} \stackrel{\text{def}}{\iff} F$ satisfies the usual definition of valuation function for pairs $\langle \tau', A \rangle$, where τ' is a subexpression of τ . Note that $\text{Free } \tau' \subseteq \text{Free } \tau$, so A is an assignment for τ' . A $\langle \tau, A \rangle$ -valuation function is finite and is therefore a set. We now show that for any term τ and assignment A for τ , there is a unique $\langle \tau, A \rangle$ -valuation function, by supposing toward a contradiction that the class C of terms τ' with $\text{Free } \tau' \subseteq \text{Dom } A$ for which there is no $\langle \tau', A \rangle$ -valuation function, is nonempty. The definition of C employs only set-quantification, so C exists. Let $\tau_0 \in C$ be of minimal complexity. One easily derives a contradiction.

Now we can define the value of τ at A as the value assigned to τ by the unique $\langle \tau, A \rangle$ -valuation function. Again, the quantification over $\langle \tau, A \rangle$ -valuation functions is set-quantification, so the valuation function exists. \square

The same argument works for quantifier-free formulas ϕ , because in this case, an assignment for ϕ is also an assignment for any subexpression of ϕ ; but this is not the case if ϕ contains quantification. For example, if $\phi = \exists v \psi$, the definition of the value of ϕ at A involves the values of ψ at assignments

$A \cup \{(v, a)\}$ for ψ , where a ranges over $|\mathfrak{S}|$. Thus, in defining the value of ϕ at A , it is not enough to quantify over $\langle \phi, A \rangle$ -valuation functions as in the proof of Theorem 2.4; instead, we must quantify over $\{\psi\}$ -valuation functions for ψ a subformula of ϕ . These are proper classes if \mathfrak{S} is a proper class, so we have no justification in \mathbf{C} for concluding that a class exists such as C in the proof of Theorem 2.4.

It is conventional to speak of valuation of formulas in terms of *satisfaction*. Given a valuation function F , the corresponding satisfaction relation S is given by

$$\langle \phi, A \rangle \in S \leftrightarrow F\langle \phi, A \rangle = 1.$$

We adapt the usual symbol for satisfaction to the representation of partial satisfaction by letting $\models^S \phi[A] \stackrel{\text{def}}{=} \langle \phi, A \rangle \in S$ when S is a $\{\phi\}$ -satisfaction relation for a structure \mathfrak{S} , and A is an \mathfrak{S} -assignment for ϕ . The preceding discussion shows that we may not be able to prove in \mathbf{C} the existence of a full satisfaction relation for a proper class structure. Indeed, it shows that we may not be able to prove the ostensibly weaker statement that for every formula ϕ there is a $\{\phi\}$ -satisfaction relation.

We do, however, have the following theorem.

Theorem 2.5. [C] *Suppose \mathfrak{S} is a ρ -structure, ψ and ψ' are ρ -formulas, and $\{\psi\}$ - and $\{\psi'\}$ -satisfaction relations exist for \mathfrak{S} . Suppose ϕ is obtained from ψ and/or ψ' by a single formula-building operation: $\neg, \vee, \wedge, \rightarrow, \leftrightarrow, \exists v, \forall v$. Then a $\{\phi\}$ -satisfaction relation exists.*

Proof. Straightforward. \square

Definition 2.6. [C] Suppose $\rho \subseteq \text{HF}$ and $n \in \omega$. Let $\Phi_n^\rho \stackrel{\text{def}}{=}$ the set of ρ -formulas of rank $< n$, i.e., $\mathcal{F}^\rho \cap V_n$.

Note that since the expression-building operations are rank-increasing, $\overline{\Phi_n^\rho} = \Phi_n^\rho$. If ρ is HF, we may use the canonical naming convention to formulate the following theorem, which may be called a metatheorem, inasmuch as it states that an infinite collection of sentences are theorems of \mathbf{C} . The theorem itself is formulated and proved in \mathbf{S} .

Theorem 2.7. [S] *Suppose ρ is a signature and ρ is HF. Suppose n is a finite ordinal. Let $\hat{\rho}$ and \hat{n} be the canonical names for ρ and n . Then $\mathbf{C} \vdash \ulcorner$ for every $(\hat{\rho})$ -structure \mathfrak{S} there exists a $\Phi_{\hat{n}}^{(\hat{\rho})}$ -satisfaction relation for $\mathfrak{S} \urcorner$.*

Proof. Let ρ be fixed. We proceed induction on the complexity of formulas using the fact that \mathbf{C} proves Theorem 2.5. \square

Definition 2.8. [C] Suppose \mathfrak{S} is a ρ -structure.

1. \mathfrak{S} is *weakly satisfactory* $\stackrel{\text{def}}{\iff}$ for every ρ -formula ϕ there exists a $\{\phi\}$ -satisfaction relation for \mathfrak{S} , equivalently, for every finite Φ there is a Φ -satisfaction relation.

2. \mathfrak{S} is *satisfactory* $\stackrel{\text{def}}{\iff}$ there exists a full satisfaction relation, i.e., an \mathcal{F}^ρ -satisfaction relation, for \mathfrak{S} , where \mathcal{F}^ρ is the class of all ρ -formulas.

Theorem 2.9. [C] *Suppose \mathfrak{S} is a structure. If \mathfrak{S} is a set then \mathfrak{S} is satisfactory, i.e., the full satisfaction relation for \mathfrak{S} exists.*

Proof. Straightforward. \square

2.3 The universal satisfaction predicate

As noted above, the existence of satisfaction relations for proper class structures is problematic, and the following definition is useful in this context.

Definition 2.10. [C] Suppose \mathfrak{S} is a ρ -structure.

1. Suppose ϕ is a ρ -formula, and A is an \mathfrak{S} -assignment for ϕ . Then $\mathfrak{S} \models^* \phi[A] \stackrel{\text{def}}{\iff}$ for every $\{\phi\}$ -satisfaction relation S for \mathfrak{S} , $\models^S \phi[A]$.
2. Suppose Θ is a ρ -theory (a class of ρ -sentences). $\mathfrak{S} \models^* \Theta \stackrel{\text{def}}{\iff}$ for every $\theta \in \Theta$, $\mathfrak{S} \models^* \theta$.

We call \models^* the *universal satisfaction predicate*. Note the use of universal, rather than existential, quantification over partial satisfaction relations in the definition of \models^* .

The following theorems are relevant. The first is the completeness theorem formulated in the essentially finitary theory C.

Theorem 2.11. [C] *Suppose Θ is a consistent theory in a countable signature. Then there is a satisfactory structure \mathfrak{S} such that $\mathfrak{S} \models \Theta$.*

Proof. The Henkin construction of a model for a consistent theory Θ proceeds by defining a complete consistent extension Θ' of Θ with witnesses, which are constants in an expanded signature \mathfrak{s}^+ . \mathfrak{S} is defined as the structure whose individuals are the \mathfrak{s}^+ -terms and whose predicates and operations are given by Θ' . Θ' also gives the full satisfaction relation for \mathfrak{S} . \square

Theorem 2.12. [C] *Suppose \mathfrak{S} is a weakly satisfactory ρ -structure, Θ is a ρ -theory, σ is a ρ -sentence, $\mathfrak{S} \models^* \Theta$, and $\Theta \vdash \sigma$. Then $\mathfrak{S} \models^* \sigma$.*

Proof. Suppose π is a proof of σ from Θ . Let Φ be the set of formulas occurring in π , and let S be a Φ -satisfaction relation for \mathfrak{S} . Each premise θ of π is then a member of Θ , so $\models^S \theta$. It is straightforward to show that $\models^S \sigma$, so $\mathfrak{S} \models^* \sigma$. \square

Corollary 2.13. [C] *Suppose \mathfrak{S} is a weakly satisfactory ρ -structure, Θ is a ρ -theory, and $\mathfrak{S} \models^* \Theta$. Then Θ is consistent.*

Proof. Suppose toward a contradiction that $\Theta \vdash \sigma \wedge \neg \sigma$ for some ρ -sentence σ . Then $\mathfrak{S} \models^* \sigma \wedge \neg \sigma$. Let S be a $\{\sigma \wedge \neg \sigma\}$ -satisfaction relation for \mathfrak{S} . Then $\models^S \sigma$ and $\models^S \neg \sigma$, but the latter implies that $\not\models^S \sigma$, a contradiction. \square

In this connection we note the following theorem.

Theorem 2.14. [C] *Suppose \mathfrak{S} is a ρ -structure and σ is a ρ -validity, i.e., $\vdash \sigma$. Then $\mathfrak{S} \models^* \sigma$.*

The proof is not quite as trivial as the theorem appears to be. To prove it in \mathbf{C} , as opposed to \mathbf{GB} , we use the existence of a deductive system for logic without identity that has the subformula property. For example, let \mathbf{LK} be the *logischer klassischer Kalkül* of Gentzen, as described in [9, Ch. 1, §2], and let \mathbf{LK}^- be the same system with the cut rule omitted. The latter has the subformula property, i.e., all formulas appearing in a proof of a sequent are (instances of) subformulas of formulas appearing in the final sequent. By the *cut-elimination theorem* (Gentzen's *Hauptsatz*), any sequent derivable in \mathbf{LK} is derivable in \mathbf{LK}^- .

To place this result in the proper perspective from the standpoint of \mathbf{C} , we digress briefly. Consider a fixed language in a signature without identity. The completeness theorem for the sequent calculus may be taken to be the following assertion:

(2.15) *If a sequent $J = (\Gamma \Rightarrow \Delta)$ is not derivable then there is an interpretation that does not satisfy J , i.e., a structure \mathfrak{S} and an \mathfrak{S} -assignment A for J such that $\mathfrak{S} \not\models J[A]$, i.e., $\mathfrak{S} \models \bigwedge \Gamma[A]$ and $\mathfrak{S} \not\models \bigvee \Delta[A]$.*

Here an *interpretation* is ordinarily understood to be a structure \mathfrak{S} and an \mathfrak{S} -assignment of all variables.

Working in \mathbf{C} , however, we must be more specific as to the meaning of ‘interpretation’. Specifically, we define *subvaluation* as in [10, §2.5]. (Briefly, this weakens the notion of partial valuation so that—for example—we may assign the value *true* to $\phi \vee \psi$ without having assigned a value to both ϕ and ψ ; it is enough to have assigned the value *true* to one of these, leaving the other unassigned; whereas in order to assign the value *false* to $\phi \vee \psi$ we must have assigned *false* to both ϕ and ψ .) We define a *J-interpretation* to be a structure \mathfrak{S} , an \mathfrak{S} -assignment of the free variables of J , and an \mathfrak{S} -subvaluation that assigns a value to each formula in J at the given assignment. The standard proof of (2.15) (e.g., the proof of Lemma 8.3 in [9]) yields the following:

(2.16) *If J is not \mathbf{LK}^- -derivable then there is a J -interpretation that does not satisfy J .*

A straightforward modification of this method yields the following:

(2.17) *If J is not \mathbf{LK} -derivable then there is a full interpretation that does not satisfy J .*

This is equivalent to the completeness theorem (2.11) stated above.

The corresponding soundness theorems for \mathbf{LK}^- and \mathbf{LK} are respectively:

(2.18) *If J is \mathbf{LK}^- -derivable then every J -interpretation satisfies J .*

(2.19) *If J is \mathbf{LK} -derivable then every full interpretation satisfies J .*

Given Infinity, the distinction between J - and full interpretations is irrelevant, as we may restrict our attention to structures that are sets, so that any J -interpretation is uniquely extendible to a full interpretation. This yields the standard model-theoretic proof of the cut-elimination theorem: if J is \mathbf{LK} -derivable then every full interpretation satisfies J , so every J -interpretation satisfies J , so J is \mathbf{LK}^- -derivable. In \mathbf{C} , of course, this proof of cut-elimination is not available, but there are effective proofs (e.g., the proof of Theorem 5.1 in [9]) that may be rendered in \mathbf{C} , and this yields the following as a theorem of \mathbf{C} :

(2.20) *A sequent J is \mathbf{LK}^- -derivable iff J is \mathbf{LK} -derivable iff every full interpretation satisfies J iff every J -interpretation satisfies J .*

The following finitary proof of Theorem 2.14, depending as it does on (2.20), is therefore another example of the value of the effective proof of cut-elimination.

Proof of Theorem 2.14. Suppose ρ is a signature without identity, σ is a ρ -sentence, σ (i.e., the sequent $0 \Rightarrow \{\sigma\}$) is \mathbf{LK} -derivable, and S is a $\{\sigma\}$ -satisfaction relation for \mathfrak{S} . By (2.20), $\models^S \sigma$. If ρ is a signature with identity, a short additional argument is necessary. \square

We will be particularly concerned with satisfaction relations for $(V; \in)$, where V is the class of all sets. Any mention of V as a structure in this article refers to $(V; \in)$ or to an essentially equivalent structure with additional defined predicates or operations.

Theorem 2.21. [S] *Suppose n is a finite ordinal, and $\theta \in V_n$ is an \mathbf{s} -sentence. Then $\mathbf{C} \vdash \ulcorner \text{The } \Phi_{(\hat{n})}^{\mathbf{s}}\text{-satisfaction relation for } V \text{ exists. Let } S \text{ be this relation. Then } (\theta) \leftrightarrow \models^S \ulcorner (\theta) \urcorner \leftrightarrow \models^S (\hat{\theta}). \urcorner$.*

Proof. By induction on n , using Theorem 2.7. \square

Theorem 2.22. [C] $(V; \in) \models^* S$.

Proof. For any individual axiom $\theta \in S$, we can prove $\ulcorner (V; \in) \models^* \ulcorner (\theta) \urcorner \urcorner$ directly. In fact, since $\mathbf{C} \vdash \theta$, (2.21) informs us that $\mathbf{C} \vdash \ulcorner (V; \in) \models^* \ulcorner (\theta) \urcorner \urcorner$. Note that the direct proof of $\ulcorner (V; \in) \models^* \ulcorner (\theta) \urcorner \urcorner$ uses θ as a premise. Thus, for example, to prove that $(V; \in) \models^* \text{Pair}$ we use the fact that pairs exist in V , arguing as follows:

Suppose S is a $\{\theta\}$ -satisfaction relation for V , where

$$\theta = \forall u, u' \exists v \forall w (w \in v \leftrightarrow w = u \vee w = u').$$

To show that $\models^S \theta$ we must show that

$$\forall x, x' \in V \exists y \in V \forall z \in V (\models^S \ulcorner [z] \in [y] \urcorner \leftrightarrow \models^S \ulcorner [z] = [x] \urcorner \vee \models^S \ulcorner [z] = [x'] \urcorner).$$

To this end, suppose $x, x' \in V$. Let $y = \{x, x'\}$. Then y is as desired.

If $\Theta \subseteq S$ is an axiom schema, we cannot rely in this way on the fact that $\mathbf{C} \vdash \Theta$ to show that $(V; \in) \models^* \Theta$, as our proof must be finite. As it happens, in these cases, it suffices to invoke a single corresponding axiom of \mathbf{C} .

Suppose, for example, that

$$\theta = \ulcorner \forall y \forall x \exists x' \forall z (z \in x' \leftrightarrow z \in x \wedge (\psi)(z, y)) \urcorner$$

is an instance of the **Comprehension** schema of **S**, where ψ is an **s**-formula with two free variables. We must show that for every $\{\theta\}$ -satisfaction relation S for V , $\models^S \theta$. Suppose, therefore, that S is a $\{\theta\}$ -satisfaction relation for V . We must show that

$$\forall y \forall x \exists x' \forall z \models^S \ulcorner z \in x' \leftrightarrow z \in x \wedge (\psi)(z, y) \urcorner.$$

Given y and x , let $x' = \{z \in x \mid \models^S \psi[z, y]\}$. Note that the existence of x' as a class follows from a single instance of the **Comprehension** schema of **C** (with parameters x, S, ψ, z, y). That x' is a set follows from the **Separation** axiom that states that the intersection of a class with a set is a set.

The **Collection** schema is handled similarly. \square

Note that Theorem 2.14 does not permit us to drop the condition of weak satisfactoriness in Theorem 2.12. In particular, despite Theorem 2.22, we have the following.

Theorem 2.23. [S] *If S is consistent then $C \not\models \ulcorner \text{for every theorem } \sigma \text{ of } S, (V; \in) \models^* \sigma \urcorner$.*

Proof. We will prove the contrapositive. Suppose $C \vdash \ulcorner \text{for every theorem } \sigma \text{ of } S, (V; \in) \models^* \sigma \urcorner$. Then the following is a proof of $\text{Con } S$ in **C**.

\ulcorner Suppose toward a contradiction that S is inconsistent. Then $S \vdash \exists u \, u \neq u$. By hypothesis, therefore, $(V; \in) \models^* \exists u \, u \neq u$. Let S be a $(\exists u \, u \neq u)$ -satisfaction relation for $(V; \in)$. Then $\models^S \exists u \, u \neq u$, which is clearly not the case. Hence, S is consistent. \urcorner

Thus, $C \vdash \ulcorner S \text{ is consistent} \urcorner$. Since **C** is a conservative extension of **S**, $S \vdash \ulcorner S \text{ is consistent} \urcorner$. By Gödel's second incompleteness theorem, S is therefore inconsistent. \square

Corollary 2.24. [S] *If S is consistent then $C \not\models \ulcorner (V; \in) \text{ is weakly satisfactory} \urcorner$.*

Proof. Immediate from (2.23) with (2.12). \square

Theorem 2.25. [S] *If S is consistent then S is not finitely axiomatizable.*

Our proof is closely modeled on that of Ryll-Nardzewski[6], but we work in **C**, rather than in **ZF**. Since **C** is a conservative extension of **S**, there is a proof in **S**. Since we cannot prove $\text{Con } S$ (assuming S is consistent) we must specifically assume it. Note that this implies that the theory $F = S + \neg \text{Infinity}$ is consistent. (Every satisfactory model of S has a substructure, viz., its V_ω , that is a satisfactory model of F .) We will show that F is not finitely axiomatizable, from which the theorem follows.

Proof. Suppose toward a contradiction that θ is a theorem of \mathbf{F} such that $\{\theta\} \vdash \mathbf{F}$. Given an existential \mathbf{s}' -formula $\psi = \exists v \phi(v, v_0, \dots, v_{m-1})$, let $R_\psi(n, n')$ be the formula

$$\text{Ord } n \wedge \text{Ord } n' \wedge \forall v_0, \dots, v_{m-1} \in V_n \left(\exists v \phi(v, v_0, \dots, v_{m-1}) \rightarrow \exists v \in V_{n'} \phi(v, v_0, \dots, v_{m-1}) \right),$$

and given a universal \mathbf{s}' -formula $\psi = \forall v \phi(v, v_0, \dots, v_{m-1})$, let $R_\psi(n, n')$ be the formula

$$\text{Ord } n \wedge \text{Ord } n' \wedge \forall v_0, \dots, v_{m-1} \in V_n \left(\forall v \in V_{n'} \phi(v, v_0, \dots, v_{m-1}) \rightarrow \forall v \phi(v, v_0, \dots, v_{m-1}) \right).$$

Clearly, for each such formula ψ ,

$$\mathbf{F} \vdash \forall_{\text{Ord}} n \exists_{\text{Ord}} n' > n R_\psi(n, n').$$

Let R be the conjunction of the formulas R_ψ for all existential and universal subformulas of θ . Then

$$\mathbf{F} \vdash \forall_{\text{Ord}} n \exists_{\text{Ord}} n' > n R(n, n').$$

Let $R'(n, k, f)$ be the \mathbf{s}' -formula $\ulcorner n$ and k are ordinals, $k > 1$, and f is a function with domain k such that $f(0) = n$ and for all $0 < l < k$, $f(l)$ is the least ordinal $k > f(l-1)$ such that $R(f(l-1), k) \urcorner$. Let $R''(n, k) = \exists! f R'(n, k, f)$. As we have just seen,

$$\mathbf{F} \vdash \forall_{\text{Ord}} n R''(n, 2).$$

It is also clear that

$$\mathbf{F} \vdash \forall_{\text{Ord}} n \forall_{\text{Ord}} k > 1 (R''(n, k) \rightarrow R''(n, k+1)).$$

Hence, by the appropriate instance of the induction schema (**Foundation**) of \mathbf{S} ,

$$\mathbf{F} \vdash \forall_{\text{Ord}} n \forall_{\text{Ord}} k > 1 R''(n, k).$$

Now let \mathbf{s}'' be the expansion of the signature \mathbf{s}' by the addition of a single constant symbol N , and let Θ be the \mathbf{s}'' -theory $\mathbf{F} \cup \{\ulcorner \text{Ord } N \urcorner\} \cup \{\ulcorner N > (\hat{n}) \urcorner \mid n \in \omega\}$. Recall that for any $x \in \mathbf{HF}$, \hat{x} is a canonical \mathbf{s}' -term denoting x . Since \mathbf{F} is consistent, so is Θ . By the completeness theorem 2.11, there is a satisfactory structure \mathfrak{S} such that $\mathfrak{S} \models \Theta$.

The ordinals of \mathfrak{S} are linearly ordered, with an initial segment isomorphic to $(\omega; <)$. We suppose for simplicity that this segment (with its order) is actually $(\omega; <)$.

Let $\bar{N} = N^{\mathfrak{S}}$ and let \bar{f} be the unique element of $|\mathfrak{S}|$ such that $\mathfrak{S} \models R'[\bar{N}, \bar{N}, \bar{f}]$. Let f be the corresponding function from the ordinals in \mathfrak{S} preceding \bar{N} into $\text{Ord}^{\mathfrak{S}}$.

Let \mathfrak{S}' be the substructure of \mathfrak{S} such that $|\mathfrak{S}'| = \bigcup_{n \in \omega} V_{f(n)}^{\mathfrak{S}}$. By construction, \mathfrak{S}' is an initial segment of \mathfrak{S} containing \bar{N} and is a $\{\theta\}$ -elementary substructure of \mathfrak{S} . Hence, $\mathfrak{S}' \models \theta$, so $\mathfrak{S}' \models F$, and

$$\mathfrak{S}' \models \exists f' R'(N, N, f').$$

Let \bar{f}' be the unique element of $|\mathfrak{S}'|$ such that

$$\mathfrak{S}' \models R'[\bar{N}, \bar{N}, \bar{f}'],$$

and let f' be the corresponding function from the ordinals in \mathfrak{S}' preceding \bar{N} into $\text{Ord}^{\mathfrak{S}'}$. Note that $\text{Dom } f' = \text{Dom } f$. Since \mathfrak{S}' is a $\{\theta\}$ -elementary substructure of \mathfrak{S} , $f' = f$. This contradicts the fact that $f \upharpoonright \omega$ is cofinal in $\text{Ord}^{\mathfrak{S}'}$. \square

2.4 Inner models, elementary embeddings, etc.

The theory of partial satisfaction outlined above provides simple and useful definitions of some of the more problematic notions in the metatheory of set theory. For example, an *inner model* may be defined as a transitive proper class M such that $(M; \in) \models^* \text{ZF}$.⁴ An elementary substructure of a ρ -structure \mathfrak{S} (which may be a proper class) may be defined as a substructure \mathfrak{S}' of \mathfrak{S} with the property that for every ρ -formula ϕ and every $\{\phi\}$ -satisfaction relation S for \mathfrak{S} , the restriction of S to \mathfrak{S}' is a $\{\phi\}$ -satisfaction relation for \mathfrak{S}' . A function $j : \mathfrak{S} \rightarrow \mathfrak{T}$ is elementary iff it is an isomorphism of \mathfrak{S} with an elementary substructure of \mathfrak{T} .

Given the simplicity and utility of these ideas, it is somewhat surprising that they have not gained greater currency in the exposition of the metatheoretical aspects of set theory, especially given that fact that it is standard practice to employ proper classes for this purpose—albeit informally as the extensions of formulas (in which quantification is necessarily restricted to sets, since only sets are actually supposed to exist). Since the definition of \models^* , etc., involves quantification over classes, the explanation may reside in an instinctive aversion to such quantification, deriving from a knowledge of the paradoxes that lurk beyond the pale.

Adherence to the convention that class variables are not to be quantified is illustrated in [8], which provides a formal definition of *inner model* by means of a particular conjunction σ of axioms of ZFC and a formula $\text{Inn}(M)$ such that

1. $\text{Inn}(M)$ is $\ulcorner (\sigma^M) \urcorner$ and (M) is a transitive proper class⁷, and
2. for every theorem θ of ZFC, $\text{ZFC} \vdash \text{Inn}(M) \rightarrow \theta^M$.

Another approach is to define an inner model as a transitive class that is almost universal and closed under Gödel operations[3, p.182]. With any of these approaches other *ad hoc* arrangements are necessary to deal with elementary embeddings of inner models and related notions.

⁴Note that according to this definition, it is a theorem of GB that V is an inner model of ZF and L is an inner model of ZFC.

In [5, Appendix X] Levy makes the exclusion of class quantification explicit in his description of the logical system P^* , which extends the language \mathcal{L}^s of pure set theory by the addition of class variables, which cannot be quantified; class terms $\{x \mid \Phi(x)\}$, where x is a set variable and Φ is any formula; and axioms that define the relations of membership and equality between sets and class terms. The stipulation that class variables are not to be quantified allows for a relatively easy syntactical proof (compared to [7]) that P^* is a conservative extension of the theory P , which is essentially the axiom of extensionality in \mathcal{L}^s .

3 Conservative extension of set theories to finitely axiomatizable class theories

Theorem 3.1. [C] *For any recursively enumerable extension Θ of S , there is a finite extension Θ' of C such that $\overline{\Theta'} \cap \mathcal{L}^s = \Theta$.*

Note that this is the strongest possible theorem along these lines, inasmuch as the deductive closure of a finite theory is necessarily recursively enumerable. Note also that it is proved in C , which does not have **Infinity**. Since C is a conservative extension of S , if we stated the straightforward translation of the theorem into \mathcal{L}^s it would be a theorem of S .

This should not be confused with the superficially similar results of Kleene[4] and Craig and Vaught[2], which produce a finitely axiomatizable conservative extension of a recursively enumerable ρ -theory Θ (which is required to have only infinite models) by introducing additional predicates with axioms asserting that they represent the ρ -language and satisfaction predicate, and that all sentences of Θ are true.

To highlight the issues surrounding the existence of satisfaction relations and the implications for provability, we first give a proof of the theorem in ZF , i.e., using **Infinity**, and we then show how to accomplish it in C .

Infinitary proof of Theorem 3.1. Suppose Θ is a recursively enumerable extension of S . Let D be an s -formula with two free variables, all of whose quantifiers are bounded, such that for all $x \in HF$, $x \in \Theta \leftrightarrow \exists y \in HF \ D(x, y)$.

Let $\Theta' = C \cup \{\theta'\}$, where $\theta' =$

$$(3.2) \quad \ulcorner \forall S \forall n \in \omega \forall x \in V_n \text{ (if } \exists y \in V_n \ (D)(x, y) \text{ and } S \text{ is the } \Phi_n^s\text{-satisfaction relation for } V, \text{ then } \models^S x) \urcorner.$$

θ' says roughly that for every sentence $x \in \Theta$, $V \models^* x$, making use of a slight modification of the universal satisfaction predicate as given by Definition 2.10. Since Theorem 3.1 does not mention satisfaction, we are free to define it as we wish within the proof.

Claim 3.3. $\overline{\Theta'} \cap \mathcal{L}^s \supseteq \overline{\Theta}$.

PROOF. Suppose $\theta \in \Theta$. Let $n \in \omega$ be such that $\theta \in V_n$ and there exists $y \in V_n$ such that $D(\theta, y)$. It is easy to show that $C \vdash \ulcorner \hat{\theta} \urcorner \in V_{(\hat{n})} \wedge (\hat{y}) \in$

$V_{(\hat{n})} \wedge (D(\hat{\theta}, \hat{y}))^\top$. Thus, $\Theta' \vdash \ulcorner$ for all S , if S is the $\Phi_{(\hat{n})}^s$ -satisfaction relation for V then $\models^S (\hat{\theta})^\top$. It follows from Theorem 2.21 that $\Theta' \vdash \theta$. $\square(3.3)$

Claim 3.4. $\overline{\Theta'} \cap \mathcal{L}^s \subseteq \overline{\Theta}$.

PROOF. Suppose $\sigma \in \overline{\Theta'} \cap \mathcal{L}^s$. We will use the completeness theorem to show that $\Theta \vdash \sigma$ by showing that σ holds in any satisfactory countable model of Θ . Suppose, therefore, that $\mathfrak{M} = (M; E)$ is a satisfactory countable model of Θ . Like all models of S , \mathfrak{M} has an initial segment that is isomorphic to $(\text{HF}; \in)$, and in the interest of efficiency we arrange that this initial segment actually is $(\text{HF}; \in)$. Since we are working in ZF , there are plenty of sets outside HF .

We now extend \mathfrak{M} to a model $\mathfrak{M}' = (M, M'; E')$ of \mathbf{C} by adding, for each subset A of M definable over \mathfrak{M} (from a parameter in M) that is not already the E -extension of a member of M , a new element whose extension is A . (These are the proper classes.) We will identify each $X \in M' \setminus M$ with $\{x \in M \mid x E' X\}$.

This is the standard construction for the infinitary proof that \mathbf{C} is a conservative extension of S . Since we are working in ZF , \mathfrak{M}' is a set, and therefore is satisfactory, so all references to satisfaction in \mathfrak{M}' may be understood in the usual way. It is easy to show that $\mathfrak{M}' \models \mathbf{C}$.

Claim 3.5. $\mathfrak{M}' \models \theta'$.

PROOF. Remember that we have arranged that HF is an initial segment of \mathfrak{M} and therefore of $\text{HF}^{\mathfrak{M}'}$. Suppose toward a contradiction that $n \in \omega^{\mathfrak{M}}$, $\theta, y \in (V_n)^{\mathfrak{M}}$, and $S \in M'$ are such that

- (3.6) $\mathfrak{M}' \models$
1. $D[\theta, y]$,
 2. $\ulcorner [S] \text{ is the } \Phi_{[n]}^s\text{-satisfaction relation for } V^\top$, and
 3. $\ulcorner \not\models^{[S]} [\theta]^\top$.

Suppose first that n is in the standard part of $\omega^{\mathfrak{M}}$. Then $n \in \omega$; $\theta, y \in V_n$; and $D(\theta, y)$; so $\theta \in \Theta$. Thus, $\mathfrak{M} \models \theta$. Also, $\{x \in M \mid \mathfrak{M}' \models \ulcorner [x] \in [S]^\top\}$ is the Φ_n^s -satisfaction relation for \mathfrak{M} . Thus, by virtue of (3.6.3), $\mathfrak{M}' \models \neg \theta$, contrary to the fact that $\mathfrak{M} \models \theta$.

Thus, n is in the nonstandard part of $\omega^{\mathfrak{M}}$. Note that $V_\omega \subseteq V_n^{\mathfrak{M}}$, and it is straightforward to show by induction within V_ω that S restricted to formulas in V_ω is the satisfaction relation for \mathfrak{M} , i.e., for any s -formula ψ and \mathfrak{M} -assignment A for ψ ,

$$\mathfrak{M}' \models \ulcorner \models^{[S]} [\psi][[A]]^\top \leftrightarrow \mathfrak{M} \models \psi[A]. \quad (3.7)$$

By virtue of the construction of \mathfrak{M}' , S is definable over \mathfrak{M} from a parameter in M . This allows us to apply the Gödel-Tarski theorem on the undefinability of truth, the proof of which we reprise for the present application.

Let $\phi \in V_\omega$ and $z \in M$ be such that ϕ is an \mathbf{s} -formula with free variables v_0, v_1, v_2 , and for all $\psi, A \in M$ such that $\mathfrak{M} \models \ulcorner [\psi] \urcorner$ is an \mathbf{s} -formula and $[A]$ is an assignment of its free variables⁷,

$$\mathfrak{M}' \models \ulcorner \models^{[S]} [\psi][[A]] \urcorner \leftrightarrow \mathfrak{M} \models \phi[\psi, A, z].$$

It is easy to obtain from ϕ an \mathbf{s} -formula ϕ' such that for all $\psi, a \in M$ such that $\mathfrak{M} \models \ulcorner [\psi] \urcorner$ is an \mathbf{s} -formula with free variables v_0, v_1 ⁷,

$$\mathfrak{M}' \models \ulcorner \models^{[S]} [\psi][[a], [z]] \urcorner \leftrightarrow \mathfrak{M} \models \phi'[\psi, a, z]. \quad (3.8)$$

Let $\psi = \neg \phi'(v_0, v_0, v_1)$. Note that $\phi', \psi \in V_\omega$, so by virtue of (3.8) and (3.7),

$$\begin{aligned} \mathfrak{M} \models \psi[\psi, z] &\leftrightarrow \mathfrak{M} \models \neg \phi'[\psi, \psi, z] \\ &\leftrightarrow \mathfrak{M}' \models \ulcorner \not\models^{[S]} [\psi][[\psi], [z]] \urcorner \\ &\leftrightarrow \neg \mathfrak{M} \models \psi[\psi, z] \\ &\leftrightarrow \mathfrak{M} \models \neg \psi[\psi, z]. \end{aligned}$$

This contradiction establishes the claim. $\square(3.5)$

Recall that we have supposed that $\sigma \in \overline{\Theta'} \cap \mathcal{L}^s$, and that \mathfrak{M} is an arbitrary satisfactory countable model of Θ . We have constructed the satisfactory structure \mathfrak{M}' with the same “sets” as \mathfrak{M} , and we have shown that $\mathfrak{M}' \models \Theta'$. Hence $\mathfrak{M}' \models \sigma$, so $\mathfrak{M} \models \sigma$. Thus, $\Theta \vdash \sigma$, i.e., $\sigma \in \overline{\Theta}$. $\square(3.4)$

Claims 3.3 and 3.4 together establish the theorem. \square

Finitary proof of Theorem 3.1. The finitary⁵ proof is identical to the preceding argument through the proof of Claim 3.3 but differs beginning with the proof of Claim 3.4. We restate the claim here.

Claim 3.9. $\overline{\Theta'} \cap \mathcal{L}^s \subseteq \overline{\Theta}$.

PROOF. Suppose $\sigma \in \overline{\Theta'} \cap \mathcal{L}^s$, and suppose toward a contradiction that $\Theta \not\models \sigma$, i.e., $\Theta \cup \{\neg \sigma\}$ is consistent. Using Theorem 2.11, let $\mathfrak{M} = (M; E)$ be a satisfactory structure such that $\mathfrak{M} \models \Theta \cup \{\neg \sigma\}$, and let T be the satisfaction relation for \mathfrak{M} .

Like all models of \mathbf{S} , \mathfrak{M} has an initial segment that is isomorphic to $(\mathbf{HF}; \in)$. Since it is possible that $V = \mathbf{HF}$, we do not suppose that \mathbf{HF} is itself an initial segment of \mathfrak{M} ; on the contrary, we arrange that $V \setminus |\mathfrak{M}|$ is a proper class, and that $M \cap M_1 = \emptyset$, where M_1 is the class of $\langle \phi, a \rangle$ such that ϕ is an \mathbf{s} -formula with two free variables v_0, v_1 ; $a \in M$; and $\{x \mid \models^T \phi[x, a]\}$ is not the E -extension of a member of M .

Let $E_1 = \{\langle x, \langle \phi, a \rangle \rangle \mid \langle \phi, a \rangle \in M_1 \wedge \models^T \phi[x, a]\}$. Let $M' = M \cup M_1$ and $E' = E \cup E_1$, and let $\mathfrak{M}' = (M, M'; E')$. \mathfrak{M}' is the canonical expansion of \mathfrak{M} to a model of \mathbf{C} ; however, whereas in \mathbf{ZF} we could let M' consist of definable subsets of M , we now let M' consist of definitions of subclasses of M . The

⁵Note that we construe ‘finitary’ broadly to include \mathbf{C} . As pointed out in the remark following the statement of the theorem, the theorem and proof could be given in \mathbf{S} .

reason, of course, is that M may be a proper class. As a consequence, distinct elements of M' may have the same E' -extension. We nevertheless define

$$=\mathfrak{M}' = \{\langle x, y \rangle \in M' \times M' \mid \forall z \in M (\langle z, x \rangle \in E' \leftrightarrow \langle z, y \rangle \in E')\}.$$

The equivalence classes of $=\mathfrak{M}'$ may be proper classes, so the reduction of \mathfrak{M}' to a standard model of logic with identity (i.e., one for which the identity predicate is interpreted as the identity relation) would be slightly more involved than usual. It could be done, but there is no need, so we don't bother.

Let Ψ be the class of \mathbf{c} -formulas with class-quantifier depth at most 2 (counting quantifier depth in terms of alternations of existential and universal quantification). Note that we impose no restriction on set quantifiers. All axioms of \mathbf{C} may be formulated as sentences in Ψ . Using the fact that T is the full satisfaction relation for \mathfrak{M} , we can easily show that the Ψ -satisfaction relation exists for \mathfrak{M}' (essentially by direct definition), and we let T' be this relation. Note that $T' \supseteq T$. It is straightforward to show that for every θ that is an axiom of \mathbf{C} or an axiom of identity, $\models^{T'} \theta$.

We now state and prove the analog of Claim 3.5 for the present situation. Note that θ' , as defined in (3.2), is in Ψ .

Claim 3.10. $\models^{T'} \theta'$.

PROOF. Suppose toward a contradiction that $\not\models^{T'} \theta'$. By the definition of T' , this implies that there exist $S \in M'$; $n \in \omega^{\mathfrak{M}}$; and $\theta, y \in (V_n)^{\mathfrak{M}}$ such that (3.6) holds, where we may substitute $\models^{T'}$ for $\models^{\mathfrak{M}'}$.

As before, we first suppose n is in the standard part of $\omega^{\mathfrak{M}}$. Let H be the standard part of $\text{HF}^{\mathfrak{M}}$, and let $\iota : H \rightarrow \text{HF}$ be the (unique) isomorphism. For notational convenience, let $\bar{x} = \iota(x)$ for $x \in H$, and extend this notation to assignment functions, so that $\bar{A}(\bar{v}) = A(v)$ for any assignment A and variable v in the sense of \mathfrak{M} . Then $\bar{n} \in \omega$; $\bar{\theta}, \bar{y} \in V_{\bar{n}}$; and $D(\bar{\theta}, \bar{y})$; so $\bar{\theta} \in \Theta$. Thus, $\mathfrak{M} \models \bar{\theta}$.

It is straightforward to show by induction on complexity that S agrees with T (the full satisfaction relation for \mathfrak{M}), i.e., $\models^{T'} \ulcorner \models^{[S]} [\phi][[A]]^\top \urcorner$ iff $\models^T \bar{\phi}[\bar{A}]$, i.e., $\mathfrak{M} \models \bar{\phi}[\bar{A}]$. Thus, by virtue of (3.6.3), $\mathfrak{M} \not\models \bar{\theta}$, contrary to the fact that $\mathfrak{M} \models \bar{\theta}$.

n is therefore in the nonstandard part of $\omega^{\mathfrak{M}}$. As before, it is straightforward to show by induction within V_ω that S restricted to formulas ψ in the standard part of \mathfrak{M} agrees with T , i.e.,

$$\models^{T'} \ulcorner \models^{[S]} [\psi][[A]]^\top \urcorner \leftrightarrow \mathfrak{M} \models \bar{\psi}[\bar{A}]. \quad (3.11)$$

As before, since S (actually, $\{x \in M \mid \langle x, S \rangle \in E'\}$) is definable over \mathfrak{M} , there exist $z \in M$ and $\phi' \in H$ such that $\bar{\phi}'$ is an \mathbf{s} -formula with free variables v_0, v_1, v_2 —and for all $\psi, a \in M$ such that $\mathfrak{M} \models \ulcorner \psi \urcorner$ is an \mathbf{s} -formula with free variables v_0, v_1 ,

$$\models^{T'} \ulcorner \models^{[S]} [\psi][[a], [z]]^\top \urcorner \leftrightarrow \mathfrak{M} \models \bar{\phi}'[\psi, a, z]. \quad (3.12)$$

Let $\psi = \neg\phi'(v_0, v_0, v_1)$ in the sense of \mathfrak{M} . Then $\psi \in H$ and $\bar{\psi} = \neg\bar{\phi}'(v_0, v_1, v_2)$, so by virtue of (3.12) and (3.11),

$$\begin{aligned} \mathfrak{M} \models \bar{\psi}[\psi, z] &\leftrightarrow \mathfrak{M} \models \neg\bar{\phi}'[\psi, \psi, z] \\ &\leftrightarrow \models^{T'} \neg \models^{[S]} [\psi][[\psi], [z]]^\neg \\ &\leftrightarrow \neg \mathfrak{M} \models \bar{\psi}[\psi, z] \\ &\leftrightarrow \mathfrak{M} \models \neg\bar{\psi}[\psi, z]. \end{aligned}$$

This contradiction establishes the claim. \square (3.10)

Recall that we have supposed that $\sigma \in \overline{\Theta'} \cap \mathcal{L}^s$, i.e., σ is an s-sentence and $\Theta' \vdash \sigma$; and we are attempting to show that $\Theta \vdash \sigma$, i.e., $\Theta \cup \{\neg\sigma\}$ is inconsistent. We have supposed toward a contradiction that \mathfrak{M} is a satisfactory structure such that $\mathfrak{M} \models \Theta \cup \{\neg\sigma\}$. We have constructed \mathfrak{M}' such that \mathfrak{M}' has the same “sets” as \mathfrak{M} , and $\models^{T'} \Theta' \cup \{\neg\sigma\}$, with the Ψ -satisfaction relation T' . To complete the proof we must derive a contradiction from the fact that $\Theta' \vdash \sigma$.

Theorem 2.13 is not available, because we have not shown that \mathfrak{M}' is weakly satisfactory. It is sufficient, however, that \mathfrak{M} is satisfactory. We make use of the methodology of the finitary proof of (2.2),⁶ which proceeds by showing how to transform a C-proof π of an s-sentence ν into an S-proof π' of ν by the systematic elimination of class variables in favor of class constants, and the elimination of the latter in favor of expressions involving s-formulas essentially serving as definitions. In this process, for each class constant C we define an appropriate s-formula ϕ_C with one free variable, and we replace each expression $\tau \in C$ by $\phi_C(\tau)$. The premises of π are axioms of C, each of which is replaced in π' by finitely many instances of axioms of S. ν is not affected by this transformation.

If we apply this procedure to a Θ' -proof of σ , we arrive at an s-proof with premises that are instances of S-axioms and sentences θ^ϕ obtained from θ' by omitting the universal quantification of S and replacing each expression $\tau \in S$ in (3.2) (which we imagine to be written out in full) by $\phi(\tau)$, where ϕ is an s-formula with one free variable. By construction, for any s-formula ϕ with one free variable, $\{x \in M \mid \mathfrak{M} \models \phi[x]\}$ is $\{x \in M \mid \models^{T'} \neg[x] \in [S]^\neg\}$ for some $S \in M'$. Since $\models^{T'}$ is the Ψ -satisfaction relation for \mathfrak{M}' , where Ψ is the class of c-formulas with class-quantifier depth at most 2, and $\models^{T'} \theta'$, it follows that $\models^{T'} \theta^\phi$. Hence $\mathfrak{M} \models \theta^\phi$. (We could also use the fact that $\models^{T'} C \cup \{\theta'\}$, and each θ^ϕ has a proof from $C \cup \{\theta'\}$ all of whose formulas are in Ψ .)

Thus, we have an s-proof of σ from premises that are true in \mathfrak{M} , from which it follows that $\mathfrak{M} \models \sigma$. This contradiction establishes Claim 3.9 and with it the theorem. \square

4 Universal satisfaction and forcing

The basic construction in the theory of forcing is that of a *generic extension* $M^{\mathbb{P}, G}$ of a transitive model M of ZF by an M -generic filter G on a partial

⁶We are paraphrasing somewhat Shoenfield’s original finitary proof of this result in [7]. A proof that is directly applicable to the present situation may be found in [10, Chap. 2].

order $\mathbb{P} \in M$, where $M^{\mathbb{P},G}$ is the structure that interprets each *forcing term* $x \in M^{\mathbb{P}}$ as x^G . M is referred to as the *ground model*. The usefulness of this construction derives from the existence of a *forcing relation* $\Vdash^{M,\mathbb{P}}$ with the following properties:

1. For any sentence σ of the *forcing language* $\mathcal{L}^{M,\mathbb{P}}$,
 1. if for every $p \in |\mathbb{P}|$ there is an M -generic filter G on \mathbb{P} with $p \in G$, then for any $p \in |\mathbb{P}|$, $p \Vdash^{M,\mathbb{P}} \sigma$ iff for every M -generic filter G on \mathbb{P} , if $p \in G$ then $M^{\mathbb{P},G} \models \sigma$; and
 2. for any M -generic filter G on \mathbb{P} , $M^{\mathbb{P},G} \models \sigma$ iff for some $p \in G$, $p \Vdash^{M,\mathbb{P}} \sigma$.
2. $\Vdash^{M,\mathbb{P}}$ is, in a suitable sense, definable over M , i.e., over the structure $(M; \in)$.

Note that as far as M is concerned, $\Vdash^{M,\mathbb{P}}$ is $\Vdash^{V,\mathbb{P}}$, where V is the class of all sets, so proper classes as ground models are an inescapable feature of the theory of forcing, with V as the paradigm. Given a partial order \mathbb{P} , we let $\Vdash^{\mathbb{P}}$ be $\Vdash^{V,\mathbb{P}}$.

When the ground model M is a proper class, the assumption that M models ZF cannot be formulated as $\ulcorner M \models \text{ZF} \urcorner$, as the full satisfaction relation for M may not exist. In the context of ZF, we may implement this assumption by positing $\ulcorner \theta^M \urcorner$ for every axiom θ of ZF (M being given by means of a defining formula). In the context of GB we have the option of formulating this assumption as $\ulcorner M \models^* \text{ZF} \urcorner$. This use of \models^* in the context of forcing raises an issue that does not arise in other applications of \models^* in set theory, which we will describe presently, and which we will settle by means of a conservative extension result, Theorem 4.9, somewhat akin to Theorem 3.1.

The issue in question arises in connection with the method of “arguing in the generic extension” $M^{\mathbb{P},G}$ to prove that a sentence of the forcing language is forced by a condition p . Actually, one does not argue in $M^{\mathbb{P},G}$ itself, but rather in a theory appropriate to it, without assuming that G actually exists, and uses the existence of the argument to infer the forcing relationship. Thus, this method is applicable in particular to $\Vdash^{\mathbb{P}}$, i.e., to $\Vdash^{V,\mathbb{P}}$, even though V -generic filters demonstrably do not exist.

Rather than working in the forcing language $\mathcal{L}^{M,\mathbb{P}}$ *per se*, we will work in a more conventional language, with a signature \mathbf{s}^* that extends the signature \mathbf{s} of set theory by the addition of a unary predicate symbol \mathbf{V} and constant symbols \mathbf{P} and \mathbf{G} .

(4.1) Let Θ be the \mathbf{s}^* -theory consisting of

1. ZF with the additional predicate symbol \mathbf{V} ;
2. $\ulcorner \mathbf{V} \text{ is transitive and contains every ordinal} \urcorner$, i.e., $\ulcorner \forall x, y (\mathbf{V}(x) \wedge y \in x \rightarrow \mathbf{V}(y)) \text{ and } \forall_{\text{Ord}} \alpha \mathbf{V}(\alpha) \urcorner$;
3. $\text{ZF}^{\mathbf{V}}$, i.e., all axioms of ZF relativized to \mathbf{V} ;
4. $\ulcorner \mathbf{V}(\mathbf{P}) \text{ and } \mathbf{P} \text{ is a partial order} \urcorner$;

5. $\ulcorner \mathbf{G}$ is a \mathbf{V} -generic filter on $\mathbf{P} \urcorner$;
6. \ulcorner every set is $x^{\mathbf{G}}$ for some $x \in V^{\mathbf{P}} \urcorner$.

The following proposition is a key element of the theory of generic extensions.

Proposition 4.2. [ZF] *For any finite subset T of Θ there is a finite subset F of ZF such that for any transitive (set) model M of F , partial order $\mathbb{P} \in M$, and M -generic filter G on \mathbb{P} , $M[G] \models T$ with $\mathbf{V}, \mathbf{P}, \mathbf{G}$ interpreted respectively as M, \mathbb{P} , and G .*

At this point, some remarks concerning the definability of the forcing relation are in order. Since $\Vdash^{M, \mathbb{P}}$ subsumes the satisfaction relation for a transitive class M , the full forcing relation—like the full satisfaction relation for M —is not definable over M , and $\Vdash^{\mathbb{P}}$ is not definable in the context of ZF. Instead, we define, for each formula ϕ , the relation

$$\{ \langle p, x_0, \dots, x_{n-1} \rangle \mid p \in |\mathbb{P}| \wedge x_0, \dots, x_{n-1} \in V^{\mathbb{P}} \wedge p \Vdash^{\mathbb{P}} \ulcorner \phi \urcorner(x_0, \dots, x_{n-1}) \urcorner \}. \quad (4.3)$$

This is, of course, a definition schema, not a single definition.⁷ In GB we have the option of defining a universal forcing predicate analogous to the universal satisfaction predicate \models^* , but this is irrelevant to the present discussion, so we defer this definition for now.

The following proposition establishes the method of “arguing in a generic extension”. The theorem is well known, so we only briefly sketch the proof.

Proposition 4.4. [S] *Suppose ψ and ϕ are s-formulas with $n + 2$ and n free variables, respectively. Suppose $\Theta \vdash$*

$$(4.5) \quad \ulcorner \text{for all } p \in |\mathbf{P}| \text{ and } x_0, \dots, x_{n-1} \in V^{\mathbf{P}}, \text{ if } (\psi^{\mathbf{V}})(\mathbf{P}, p, x_0, \dots, x_{n-1}) \text{ then } p \in \mathbf{G} \rightarrow (\phi)(x_0^{\mathbf{G}}, \dots, x_{n-1}^{\mathbf{G}}) \urcorner.$$

Then $\mathbf{ZF} \vdash$

$$(4.6) \quad \ulcorner \text{if } \mathbb{P} \text{ is a partial order, } p \in |\mathbb{P}|, x_0, \dots, x_{n-1} \in V^{\mathbb{P}}, \text{ and } (\psi)(\mathbb{P}, p, x_0, \dots, x_{n-1}), \text{ then } p \Vdash^{\mathbb{P}} \ulcorner \phi \urcorner(x_0, \dots, x_{n-1}) \urcorner.$$

Proof. Let ψ and ϕ be given, and let Θ_0 be a finite subset of Θ such that Θ_0 proves (4.5). Let F be a finite subset of ZF such that for any transitive model M of F , and any partial order $\mathbb{P} \in M$,

1. M correctly defines $\Vdash^{M, \mathbb{P}} \phi$; and

⁷Alternatively, we may extend ZF by the addition of a new predicate symbol $\ulcorner \Vdash \urcorner$, with axioms that correspond to the usual recursive definition of the forcing relation. Note that these axioms allow us to generate a definition for the relation (4.3) for any given ϕ , but this definition has quantifier depth that increases with that of ϕ , and the axioms do not yield a definition of $\Vdash^{\mathbb{P}}$ in its entirety. Note also that, since $\ulcorner \Vdash \urcorner$ is not introduced by definition, we must explicitly extend the axiom schemas of ZF to formulas that incorporate the new symbol. It is not hard to show that this theory is a conservative extension of ZF, so it is largely immaterial which approach we use to the description of forcing over V , but in this article we will deal with ZF unmodified.

2. for any M -generic filter G on \mathbb{P} , $M[G] \models \Theta_0$ with $\mathbf{V}, \mathbf{P}, \mathbf{G}$ interpreted respectively as M, \mathbb{P} , and G .

We now sketch a proof of (4.6) in ZF (without being too fussy about use vs. mention). We begin by supposing toward a contradiction that it is not the case. We use a reflection argument, followed by the transitive collapse of a countable elementary substructure, to obtain a countable transitive model M of F , with a partial order $\mathbb{P} \in M$, $p \in |\mathbb{P}|$, and $x_0, \dots, x_{n-1} \in M^{\mathbb{P}}$, such that $M \models \psi[\mathbb{P}, p, x_0, \dots, x_{n-1}]$, and $p \Vdash^{M, \mathbb{P}} \phi(x_0, \dots, x_{n-1})$. We let G be an M -generic filter on \mathbb{P} such that $p \in G$ and $M[G] \not\models \phi[x_0^G, \dots, x_{n-1}^G]$. Since $M \models F$, $M[G] \models \Theta_0$. It follows that $M[G] \models \phi[x_0^G, \dots, x_{n-1}^G]$, a contradiction.

In effect, we justify the method of arguing in a hypothetical generic extension of V by arguing in an actual generic extension of a countable transitive model of a finite fragment of ZF. \square

Note that Θ implements the hypothesis that \mathbf{V} is a model of ZF by positing each axiom of ZF relativized to \mathbf{V} . In a pure set theory we have no other option, as ZF is not finitely axiomatizable, and proper classes do not exist. In GB we may use the satisfaction predicate \models^* defined in (2.10) to implement the hypothesis that a proper class M is a model of ZF as the single sentence $\ulcorner M \models^* \text{ZF} \urcorner$.

The use of \models^* can play the same simplifying role in the exposition of the theory of forcing as in other areas of set theory that deal with proper class models, but in the case of forcing, the following question arises:

Let \mathbf{c}^* be the signature \mathbf{c} with additional constants \mathbf{V} , \mathbf{P} , and \mathbf{G} ; we also treat \mathbf{V} as a unary predicate in the usual way.

(4.7) *Let Θ' be the \mathbf{c}^* -theory which is Θ with the following changes:*

- 1'. GB.
- 3'. $\ulcorner \mathbf{V} \models^* \text{ZF} \urcorner$.

When “arguing in a generic extension”, we will naturally reason from Θ' rather than Θ . The question is whether Proposition 4.4 applies with Θ' in place of Θ .

In the absence of an affirmative answer to this question, the usefulness of \models^* is much diminished, as one must maintain a parallel development of forcing without \models^* to use when deriving forcing relations by “arguing in a generic extension”. Thus, the following theorem is a great convenience.

Theorem 4.8. [S] *Suppose ψ and ϕ are \mathbf{s} -formulas with $n+2$ and n free variables, respectively. Suppose $\Theta' \vdash$*

\ulcorner for all $p \in |\mathbb{P}|$ and $x_0, \dots, x_{n-1} \in V^{\mathbb{P}}$, if $(\psi^{\mathbf{V}})(\mathbf{P}, p, x_0, \dots, x_{n-1})$ then $p \in \mathbf{G} \rightarrow (\phi)(x_0^{\mathbf{G}}, \dots, x_{n-1}^{\mathbf{G}}) \urcorner$.

Then $\text{ZF} \vdash$

\ulcorner if \mathbb{P} is a partial order, $p \in |\mathbb{P}|$, $x_0, \dots, x_{n-1} \in V^{\mathbb{P}}$, and $(\psi)(\mathbb{P}, p, x_0, \dots, x_{n-1})$, then $p \Vdash^{\mathbb{P}} \ulcorner (\phi)(x_0, \dots, x_{n-1}) \urcorner \urcorner$.

Proof. This follows immediately from Theorem 4.4 and the following theorem (4.9). \square

Theorem 4.9. *[S] Θ' is a conservative extension of Θ in the sense that for any s^* -sentence σ , if $\Theta' \vdash \sigma$ then $\Theta \vdash \sigma$.*

Proof. We will carry out the proof in \mathbf{C} . Since the statement of the theorem is an s -sentence and \mathbf{C} is a conservative extension of \mathbf{S} , the theorem follows from \mathbf{S} .

We begin as in the finitary proof of Theorem 3.1. Thus, suppose $\Theta' \vdash \sigma$ and suppose toward a contradiction that $\Theta \not\vdash \sigma$. Let $\mathfrak{M} = (M; \in^{\mathfrak{M}}, M_0, \mathbb{P}, G)$ be a satisfactory structure such that $\mathfrak{M} \models \Theta \cup \{\neg\sigma\}$, where $M_0 = \mathbf{V}^{\mathfrak{M}}$, $\mathbb{P} = \mathbf{P}^{\mathfrak{M}}$ and $G = \mathbf{G}^{\mathfrak{M}}$. Let T be the (full) satisfaction relation for \mathfrak{M} . Thus, $\models^T \Theta \cup \{\neg\sigma\}$. Let s^V be the expansion of the signature s by the addition of the unary predicate symbol V (without the constant symbols \mathbf{P} and \mathbf{G} of s^*). Extend \mathfrak{M} to a c^* -structure \mathfrak{M}' as before, by adding “proper classes” definable over \mathfrak{M} .

(4.10) *Clearly, each added class is defined by an s^V -formula from a parameter in M , which may incorporate \mathbb{P} and/or G .⁸*

Let Ψ be the class of c^* -formulas with class-quantifier depth at most 2, and let T' be the Ψ -satisfaction relation for \mathfrak{M}' .

Claim 4.11. $\models^{T'} \Theta'$.

PROOF. It is straightforward to check that $\models^{T'} \theta$ for all $\theta \in \Theta'$ other than $\ulcorner V \models^* ZF \urcorner$. To show that $\models^{T'} \ulcorner V \models^* ZF \urcorner$, suppose toward a contradiction that it does not. Note that $\ulcorner V \models^* ZF \urcorner$ is \ulcorner for every $\theta \in ZF$, for every $\{\theta\}$ -satisfaction relation S for V , $\langle \theta, 0 \rangle \in S \urcorner$, so it has class-quantifier depth 1. Hence $\models^{T'} \ulcorner V \models^* ZF \urcorner$, so there exist $\theta \in M$ and $S \in M'$ such that $\models^T \ulcorner \theta \in ZF \urcorner$ and $\models^{T'} \ulcorner [S] \text{ is the } \{\theta\}\text{-satisfaction relation for } V, \text{ and } \not\models^{[S]} [\theta] \urcorner$.

At this point the proof of Theorem 3.1 bifurcated according to whether θ is in the standard or the nonstandard part of \mathfrak{M} . In the latter event we obtained a contradiction from the fact that S would include the full satisfaction relation for \mathfrak{M} , which cannot be definable over \mathfrak{M} . That depended on Definition 3.2 of θ' in terms of Φ_n^s -satisfaction relations: if $n \in M$ is nonstandard then Φ_n^s contains every standard s -formula. For this method (with $(V; \in)$ and θ instead of $(M; \in)$ and θ') to be applicable in the present case, we would have to reformulate the universal satisfaction predicate so that $\mathfrak{S} \models^* \phi[A]$ iff for every Φ_ϕ -satisfaction relation S for \mathfrak{S} , $\langle \phi, A \rangle \in S$, where Φ_ϕ is defined so that if ϕ is nonstandard then Φ_ϕ contains every standard formula. For example, we could let Φ_ϕ consist of all formulas with complexity not greater than that of ϕ in some appropriate sense, rather than letting Φ_ϕ consist of all subformulas of ϕ , as we have done. In the proof of Theorem 3.1 it was legitimate to define Φ_ϕ however we wished, as the theorem does not mention satisfaction. In the present case, such an

⁸For notational simplicity, any formula requiring $n > 1$ parameters is replaced by a formula with a single parameter, which is an n -sequence.

alteration would be inelegant, to say the least—and it is unnecessary, since we may proceed as follows.

As before, let H be the standard part of $\text{HF}^{\mathfrak{M}}$. Note that H is also the standard part of $\text{HF}^{\mathfrak{M}_0}$, where \mathfrak{M}_0 is the substructure of \mathfrak{M} corresponding to M_0 . Let $x \mapsto \bar{x}$ be the isomorphism of H with HF . To simplify the notation, suppose that $(\mathcal{L}^{\mathfrak{s}*})^{\mathfrak{M}}$ is $\mathcal{L}^{\mathfrak{s}*}$, so $\bar{\epsilon} = \epsilon$ for any \mathfrak{s}^* -expression ϵ .

It is easy to show (as in Theorem 2.21) that if $\theta \in H$ then $\models^{T'} \ulcorner \models^{[S]} [\theta] \urcorner$ iff $\models^T \theta^{\vee}$, so $\models^{T'} \ulcorner \models^{[S]} [\theta] \urcorner$, since $\models^T \text{ZF}^{\vee}$ by hypothesis. Thus, θ is in the nonstandard part of $\text{ZF}^{\mathfrak{M}}$, which means that it is an instance of one of the axiom schemas for a nonstandard formula.

We will suppose that the schemas are **Collection** and **Comprehension**. (Foundation may be formulated in **ZF** as the statement that all nonempty sets have an \in -minimal element; it need not be formulated as a schema.) A sufficiently general version of **Collection** is

$$\ulcorner \forall y \forall x \exists_{\text{Ord}} \alpha \forall z \in x (\exists_{\text{Ord}} \beta (\psi)(z, \beta, y) \rightarrow \exists \beta < \alpha (\psi)(z, \beta, y)) \urcorner,$$

where ψ is an \mathfrak{s} -formula with three free variables. Suppose θ is the above instance of **Collection**. We will derive a contradiction by showing that $\models^{T'} \ulcorner \models^{[S]} [\theta] \urcorner$.

Since $\models^{T'} \ulcorner [S] \urcorner$ is the $\{[\theta]\}$ -satisfaction relation for \mathbf{V}^{\ulcorner} and $\models^T \ulcorner [\psi] \urcorner$ is a subformula of $[\theta]^{\ulcorner}$, it suffices to show that

$$\begin{aligned} \models^{T'} \ulcorner \forall y, x \in \mathbf{V} \exists_{\text{Ord}} \alpha \forall z \in x (\exists_{\text{Ord}} \beta \models^{[S]} [\psi][z, \beta, y] \\ \rightarrow \exists_{\text{Ord}} \beta < \alpha \models^{[S]} [\psi][z, \beta, y]) \urcorner. \end{aligned}$$

This is an instance of the **Collection** schema of **GB** and follows from the fact that $\models^{T'}$ **Collection**.

The case that θ is an instance of **Comprehension** is not so simple. Suppose

$$\theta = \ulcorner \forall y \forall x \exists x' \forall z (z \in x' \leftrightarrow z \in x \wedge (\psi)(z, y)) \urcorner,$$

where ψ is an \mathfrak{s} -formula with two free variables. Given $y, x \in M_0$, we must show that there exists $x' \in M_0$ such that for all $z \in M_0$, $z \in^{\mathfrak{M}} x'$ iff $z \in^{\mathfrak{M}} x$ and $\models^{T'} \ulcorner \models^{[S]} [\psi][z, y] \urcorner$.

By construction, S represents a subclass of M definable over \mathfrak{M} by an \mathfrak{s}^{\vee} -formula ϕ from a parameter in M , which is a^G for some $a \in M_0$, i.e.,

$$\models^{T'} \ulcorner \models^{[S]} [\psi][[A]]^{\ulcorner} \leftrightarrow \langle \psi, A \rangle \in^{\mathfrak{M}'} S \leftrightarrow \models^T \phi[\psi, A, a^G] \urcorner.$$

Thus, given $y, x \in M_0$,

(4.12) *we must show that there exists $x' \in M_0$ such that for all $z \in M_0$, $z \in^{\mathfrak{M}} x'$ iff $z \in^{\mathfrak{M}} x$ and $\models^T \phi[\psi, A, a^G]$, where A is the assignment of z and y to the free variables of ψ .*

Let u, v be new variables, and let ϕ' be the \mathfrak{s}^{\vee} -formula with free variables u, v , obtained from $\ulcorner S \urcorner$ is the $\{(u)\}$ -satisfaction relation for \mathbf{V}^{\ulcorner} by replacing each subformula of the form $\ulcorner \langle \psi, A \rangle \in S^{\ulcorner}$ by $\phi(\psi, A, v)$. Without belaboring the issue, suffice it to say that ϕ' is a conjunction of formulas such as

1. $\ulcorner (u) \text{ is an } \mathfrak{s}^*\text{-formula} \urcorner$;
2. $\ulcorner \text{for any subformulas } \psi \text{ and } \psi' \text{ of } (u) \text{ and } \mathbf{V}\text{-assignment } A \text{ for } \psi, \text{ if } \psi = \neg \psi', \text{ then } (\phi)(\psi, A, v) \text{ iff } \neg (\phi)(\psi', A, v) \urcorner$ (with similar formulas for the other propositional connectives); and
3. $\ulcorner \text{for any subformulas } \psi \text{ and } \psi' \text{ of } (u), \text{ variable } w, \text{ and } \mathbf{V}\text{-assignment } A \text{ for } \psi, \text{ if } \psi = \exists w \psi', \text{ then } (\phi)(\psi, A, v) \text{ iff for some } x \text{ such that } \mathbf{V}(x), (\phi)(\psi', A \cup \{(w, x)\}, v) \urcorner$ (with a similar formula for the universal quantifier).

We now have

$$\models^T \phi'[\theta, a^G].$$

Since T is the full satisfaction relation for \mathfrak{M} and $\models^T \Theta$, any deduction from Θ holds in T . We will therefore argue in Θ as follows.

\ulcorner Suppose $(\phi'(\theta, a^G))$. Let $p \in G$ be such that

$$p \Vdash \ulcorner (\phi'(\check{\theta}, a)) \urcorner.$$

Claim 4.13. *For every subformula ψ of θ and every \mathbf{V} -assignment A for ψ , p decides $\ulcorner (\phi(\check{\psi}, \check{A}, a)) \urcorner$, i.e., either*

1. $p \Vdash \ulcorner (\phi(\check{\psi}, \check{A}, a)) \urcorner$, or
2. $p \Vdash \ulcorner (\neg \phi(\check{\psi}, \check{A}, a)) \urcorner$.

PROOF. Suppose not. Let ψ be a counterexample of minimal complexity. By way of illustration, suppose $\psi = \exists w \psi'$, and suppose A is a \mathbf{V} -assignment for ψ . For any $x \in \mathbf{V}$, let $A^x = A \cup \{(w, x)\}$, the extension of A that assigns x to w . By hypothesis, for any $x \in \mathbf{V}$, p decides $\ulcorner (\phi(\check{\psi}', \check{A}^x, a)) \urcorner$. Recall that p forces $\ulcorner (\phi'(\check{\theta}, a)) \urcorner$, which says that $\ulcorner (\phi) \urcorner$ defines the $\{\theta\}$ -satisfaction relation for \mathbf{V} from the parameter a^G .

Suppose $p \not\Vdash \ulcorner (\phi(\check{\psi}, \check{A}, a)) \urcorner$. Then for all $x \in \mathbf{V}$, $p \not\Vdash \ulcorner (\phi(\check{\psi}', \check{A}^x, a)) \urcorner$. Thus, for all $x \in \mathbf{V}$, since p decides $\ulcorner (\phi(\check{\psi}', \check{A}^x, a)) \urcorner$, $p \Vdash \ulcorner (\neg \phi(\check{\psi}', \check{A}^x, a)) \urcorner$. Hence, $p \Vdash \ulcorner (\neg \phi(\check{\psi}, \check{A}, a)) \urcorner$.

The other recursive clauses in the definition of satisfaction are handled similarly, and the atomic formulas are easily dealt with. $\square(4.13)$

Let x' be the set of $z \in x$ such that $p \Vdash \ulcorner (\phi(\check{\psi}, \check{A}, a)) \urcorner$, where A is the assignment of z and y to the free variables of ψ . Then $x' \in \mathbf{V}$ by virtue of **Comprehension**^V. Given $z \in \mathbf{V}$, let A be the assignment of z and y to the free variables of ψ . If $z \in x'$ then $p \Vdash \ulcorner (\phi(\check{\psi}, \check{A}, a)) \urcorner$, so $(\phi)(\psi, A, a^G)$, since $p \in G$. On the other hand, if $z \notin x'$ then $p \not\Vdash \ulcorner (\phi(\check{\psi}, \check{A}, a)) \urcorner$, so by the claim, $p \Vdash \ulcorner (\neg \phi(\check{\psi}, \check{A}, a)) \urcorner$, whence $\neg (\phi)(\psi, A, a^G)$, since $p \in G$.

As noted above, the existence of this argument in Θ shows that there exists $x' \in M_0$ as required by (4.12), and this completes the proof that $\models^{T'} \theta$. $\square(4.11)$

We now know that $\models^{T'} \Theta'$ and $\models^{T'} \neg \sigma$, where T' is the Ψ -satisfaction relation for \mathfrak{M}' , Ψ being the class of \mathfrak{c}^* -formulas with class-quantifier depth at most 2. We now wish to derive a contradiction from the assumption that $\Theta' \vdash \sigma$.

As in the proof of Theorem 3.1, if we had **Infinity** we could arrange that \mathfrak{M}' be a set and take T' to be the full satisfaction relation for \mathfrak{M}' , from which the desired contradiction would follow at once. It would not be inappropriate to finish this way, as the theorem is only of interest in the context of forcing, which is only of interest in the context of **Infinity**; however, an argument can be made that if a finitary theorem has a finitary proof, one should be given, and we oblige.

To complete the proof in **C** we proceed as in finitary proof of Theorem 3.1, showing that any proof π of σ from Θ' may be replaced by a proof π' of σ from Θ . Let $\theta' = \ulcorner \forall \models^* \text{ZF} \urcorner$, i.e.,

$$(4.14) \quad \ulcorner \text{for all } S, \text{ for all } x \in \text{ZF}, \text{ if } S \text{ is an } \{x\}\text{-satisfaction relation for } \forall, \text{ then } \langle x, 0 \rangle \in S \urcorner.$$

As before, we eliminate class variables in favor of class constants, and then replace each expression $\tau \in C$ by $\phi_C(\tau)$, where ϕ is an appropriate “definition” of C . Each axiom of **GB** used as a premise in π is replaced in π' by finitely many instances of axioms of **ZF**.

The premise θ' is replaced by finitely many sentences θ^ϕ obtained from θ' by omitting the universal quantification of S and replacing each expression $\tau \in S$ in (4.14) by $\phi(\tau)$, where ϕ is a formula with one free variable v . As before, since each such ϕ defines an element of M' and $\models^{T'} \theta'$, it follows that $\models^{T'} \theta^\phi$. Hence $\mathfrak{M} \models \theta^\phi$.

Thus, we have an s^* -proof of σ from premises that are true in \mathfrak{M} , so $\mathfrak{M} \models \sigma$, contradicting our assumption that $\mathfrak{M} \models \neg \sigma$. \square

5 The universal forcing relation

We conclude by giving the promised definition of the universal forcing relation and valuation function. We leave it to the reader to supply the definitions of ‘ $\Phi^{M, \mathbb{P}}$ -forcing relation’ and ‘ $\Phi^{M, \mathfrak{A}}$ -valuation function’, where Φ is a class of s^V -formulas. (These will cover all sentences obtained from subformulas of members of Φ by substitution of elements of $M^{\mathbb{P}}$ or $M^{\mathfrak{A}}$, respectively, for their free variables.)

Definition 5.1. [GB] Suppose M is a transitive model of **ZF**, \mathbb{P} is a partial order in M , and \mathfrak{A} is an M -complete boolean algebra in M .

1. Suppose ϕ is an $\mathcal{L}^{M, \mathbb{P}}$ -sentence and $p \in |\mathbb{P}|$. Then $p \Vdash^{*M, \mathbb{P}} \phi \stackrel{\text{def}}{\iff}$ for every $\{\phi\}^{M, \mathbb{P}}$ -forcing relation F , $p F \phi$.
2. Suppose ϕ is an $\mathcal{L}^{M, \mathfrak{A}}$ -sentence. If there exists a $\{\phi\}^{M, \mathfrak{A}}$ -valuation function F then $\llbracket \phi \rrbracket^{*M, \mathfrak{A}} \stackrel{\text{def}}{=} F \phi$; otherwise, $\llbracket \phi \rrbracket^{*M, \mathfrak{A}} \stackrel{\text{def}}{=} \mathbf{1}$.⁹

⁹ $\mathbf{1}$ is the correct value for $\llbracket \phi \rrbracket^{*M, \mathfrak{A}}$ if no $\{\phi\}^{M, \mathfrak{A}}$ -valuation function exists, because in this case no $\{\phi\}^{M, \mathbb{P}}$ -forcing relation exists, so $\{p \mid p \Vdash^{*M, \mathbb{P}} \phi\} = |\mathbb{P}|$, which corresponds to boolean value $\mathbf{1}$.

These definitions reduce to the usual ones when M is a set, and when M is a proper class they permit the development of the theory of forcing in the usual way.

As we have noted above, this universal forcing predicate is not involved in any of the considerations of the preceding section. If we wished, of course, we could reformulate the conclusion of Theorem 4.8 to be that $\text{GB} \vdash$

$\ulcorner \text{if } \mathbb{P} \text{ is a partial order, } p \in |\mathbb{P}|, x_0, \dots, x_{n-1} \in V^{\mathbb{P}}, \text{ and } (\psi)(\mathbb{P}, p, x_0, \dots, x_{n-1}), \text{ then } p \Vdash^{*\mathbb{P}} \ulcorner \phi(x_0, \dots, x_{n-1}) \urcorner \urcorner,$

which would serve the same practical purpose, and would be appropriate in an exposition of the theory of forcing in GB using \Vdash^* .

References

- [1] Paul Bernays. A system of axiomatic set theory—Part I. *Journal of Symbolic Logic*, 2(1):65–77, 1937.
- [2] W. Craig and R. L. Vaught. Finite axiomatizability using additional predicates. *Journal of Symbolic Logic*, 23(3):289–308, 1958.
- [3] Thomas J. Jech. *Set Theory*. Springer Monographs in Mathematics. Springer-Verlag, New York, third edition, 2003.
- [4] Stephen C. Kleene. Finite axiomatizability of theories in the predicate calculus using additional predicate symbols. *Memoirs of the American Mathematical Society*, (10):27–68, 1952.
- [5] Azriel Levy. *Basic Set Theory*. Springer-Verlag, New York, first edition, 1979.
- [6] Czesław Ryll-Nardzewski. The role of the axiom of induction in elementary arithmetic. *Fundamenta Mathematicae*, 39:239–263, 1952.
- [7] Joseph R. Shoenfield. A relative consistency proof. *Journal of Symbolic Logic*, 19(1):21–28, 1954.
- [8] Robert M. Solovay, W. N. Reinhardt, and A. Kanamori. Strong axioms of infinity and elementary embeddings. *Annals of Mathematical Logic*, 13(1):73–116, 1978.
- [9] Gaisi Takeuti. *Proof Theory*, volume 81 of *Studies in Logic and the Foundations of Mathematics*. Elsevier, New York, 1975.
- [10] Robert A. Van Wesep. **Foundations of Mathematics: A Generalist’s Guide**.